



FACULTÉ DE DROIT
ÉCONOMIE & GESTION

Sujet d'examen - 1^{ère} session
Semestres 1-3-5
Année universitaire 2018-2019

Intitulé de l'épreuve :	Analyse de données qualitatives
Nom de l'enseignant :	Nils Berglund
Mention / Spécialité / Parcours :	M1 ESA
Année :	2018-2019
Durée de l'épreuve :	2 heures
Documents autorisés :	Résumé manuscrit de 4 pages A4
Matériels autorisés :	Calculatrice non programmable
	19/12/2018

P1/7

SUJET

Les téléphones portables doivent être éteints durant l'examen.
Il sera tenu compte de la qualité de la rédaction.
Les points sont donnés à titre indicatif.

Problème 1 [4 points]

On considère la matrice

$$A = \begin{bmatrix} 17 & -6 \\ -6 & 8 \end{bmatrix}.$$

1. Déterminer les valeurs propres de A .
2. Déterminer, pour chaque valeur propre de A , un vecteur propre de norme 1.
3. Soit \mathcal{E} l'ellipse d'équation

$$[x \ y] A \begin{bmatrix} x \\ y \end{bmatrix} = 100 \quad (1)$$

Que vaut son demi-grand axe ? Son demi-petit axe ?

4. Esquisser l'ellipse \mathcal{E} dans le plan (x, y) .

Problème 2 [5 points]

On considère le tableau de contingence suivant. Ici a, b, c et d représentent quatre modalités d'une première variable qualitative, alors que A, B, C et D représentent quatre modalités d'une seconde variable qualitative.

	a	b	c	d	
A	10	10	30	50	100
B	50	0	10	40	100
C	20	20	60	100	200
D	20	20	0	60	100
	100	50	100	250	500

1. Définissez la notion d'équivalence distributionnelle. Utilisez cette notion afin de simplifier le tableau ci-dessus.
2. Calculez le tableau théorique en cas d'indépendance correspondant au tableau simplifié.
3. Effectuez un test d'indépendance du chi-deux du tableau simplifié.
4. Le tableau ci-dessous donne les valeurs critiques de variables du chi-deux pour différentes valeurs du nombre de degrés de liberté et du seuil. Peut-on rejeter l'hypothèse d'indépendance au seuil de 95% ?

Degrés de liberté	90%	95%	97.5%	99%	99.9%
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588
11	17.275	19.675	21.920	24.725	31.264
12	18.549	21.026	23.337	26.217	32.910
13	19.812	22.362	24.736	27.688	34.528
14	21.064	23.685	26.119	29.141	36.123
15	22.307	24.996	27.488	30.578	37.697
16	23.542	26.296	28.845	32.000	39.252

Problème 3 [6 points]

Une enquête auprès de 1565 employés dans une compagnie minière, vivant en couple mais n'ayant pas d'enfant, leur demandait la raison principale pour laquelle ils n'ont pas eu d'enfants.

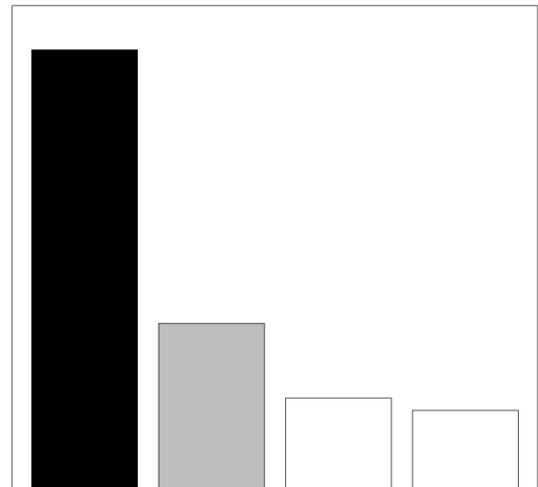
Les résultats croisant la raison invoquée et le niveau de qualification de l'employé sont indiqués dans le tableau de contingence ci-dessous.

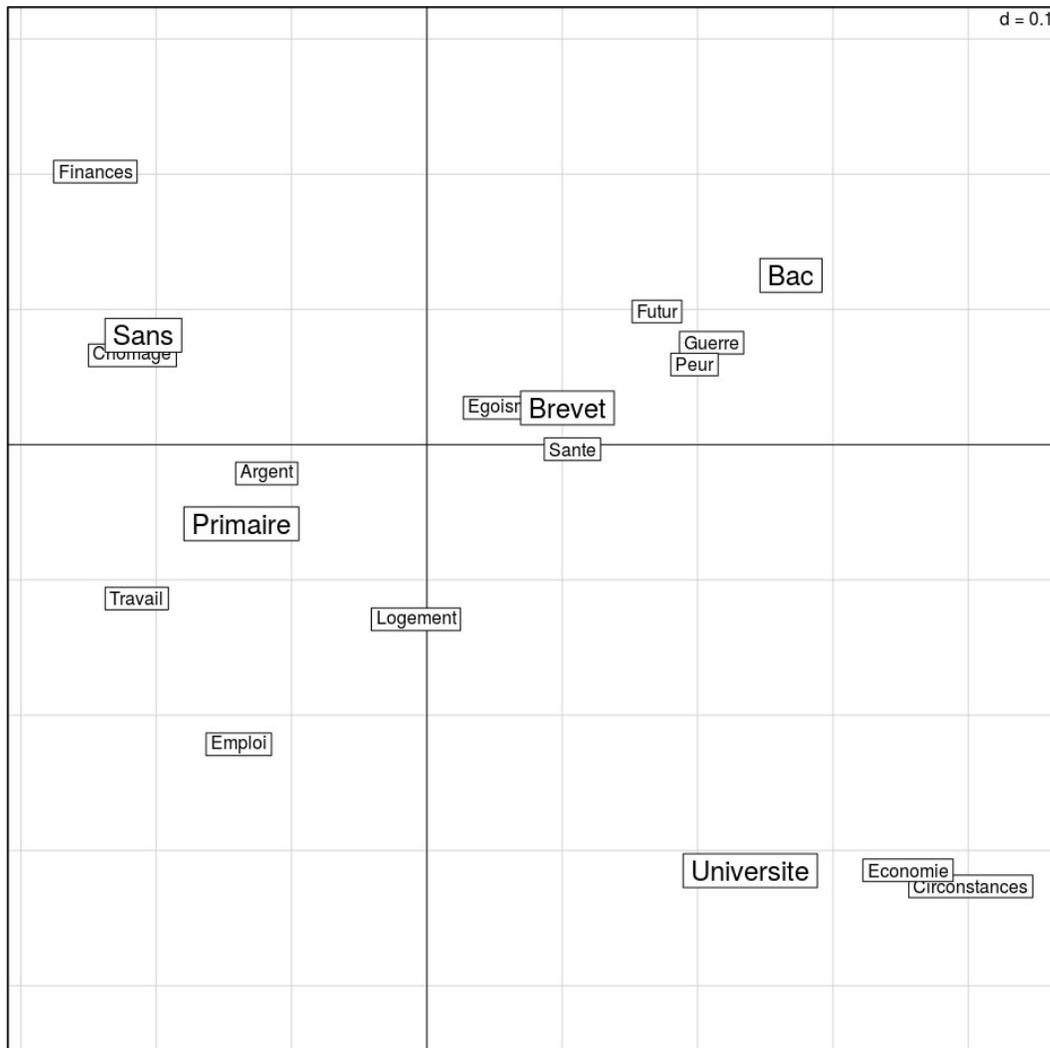
Motif \ Qualification	Sans	Ecole primaire	Brevet	Bac	Universite	Total
Argent	51	64	32	29	17	193
Futur	53	90	78	75	22	318
Chomage	71	111	50	40	11	283
Circonstances	1	7	5	5	4	22
Economie	7	13	12	11	11	54
Egoisme	21	37	14	26	9	107
Emploi	12	35	19	6	7	79
Finances	10	7	7	3	1	28
Guerre	4	7	7	6	2	26
Logement	8	22	7	10	5	52
Peur	25	45	38	38	13	159
Sante	18	27	20	19	9	93
Travail	35	61	29	14	12	151
Total	316	526	318	282	123	1565

Voici les résultats d'une AFC faite sur ces données.

```

> afcin$TOT
      inertia      cum      cum(%)
Ax1 0.034887305 0.03488730 56.42688
Ax2 0.013241258 0.04812856 77.84335
Ax3 0.007338307 0.05546687 89.71235
Ax4 0.006360590 0.06182746 100.00000
    
```





Contributions absolues et relatives des lignes (multipliées par 10000):

> afcin\$row.abs			> afcin\$row.rel		
	Axis1	Axis2		Axis1	Axis2
Argent	495	42	Argent	-4337	-139
Futur	1683	1487	Futur	6950	2331
Chomage	2454	596	Chomage	-8906	821
Circonstances	650	1131	Circonstances	5916	-3906
Economie	1249	2580	Economie	4918	-3854
Egoisme	62	38	Egoisme	658	152
Emploi	279	1870	Emploi	-1681	-4272
Finances	307	550	Finances	-2924	1986
Guerre	210	71	Guerre	7322	940
Logement	0	418	Logement	-10	-2695
Peur	1138	270	Peur	8903	801
Sante	196	0	Sante	7864	-7
Travail	1270	941	Travail	-7506	-2111

Contributions absolues et relatives des colonnes (multipliées par 10000) :

```
> afcin$col.abs           > afcin$col.rel
      Axis1 Axis2           Axis1 Axis2
Sans      2539   998   Sans      -6731  1004
Primaire  1808   868   Primaire  -6312 -1150
Brevet    627   113   Brevet    2915   199
Bac       3737  2132   Bac       7463  1615
Universite 1285  5887   Universite 3278 -5697
```

1. Expliquer les lignes suivantes fournies par le logiciel R:

```
Number of cases in table: 1565
```

```
Number of factors: 2
```

```
Test for independence of all factors:
```

```
Chisq = 96.76, df = 48, p-value = 3.894e-05
```

2. Discuter les valeurs des inerties. Une étude restreinte aux deux premiers axes factoriels vous semble-t-elle pertinente ?
3. Quelles sont les modalités mal représentées sur le plan factoriel ?
4. Quelles sont les modalités contribuant fortement au premier axe factoriel ? Interpréter cet axe.
5. Quelles sont les modalités contribuant fortement au second axe factoriel ?
6. Quelle différences y a-t-il entre la répartition des raisons invoquées par les universitaires et par les personnes sans qualification ?

Problème 4 [5 points]

Un sondage fait auprès de 20 personnes porte sur leur classe d'âge, leur préférence de restaurant parmi fast food, restaurant indien, et restaurant gastronomique (étoilé), ainsi que le budget pour un repas au restaurant.

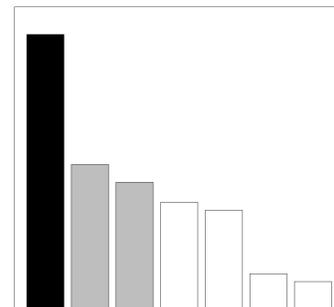
	Age	Restaurant	Budget
1	Moins de 20 ans	Fast Food	Moins de 20 euros
2	Moins de 20 ans	Fast Food	Moins de 20 ans
3	Moins de 20 ans	Fast Food	20 à 40 euros
4	Moins de 20 ans	Indien	Moins de 20 euros
5	21 à 40 ans	Fast Food	Moins de 20 euros
6	21 à 40 ans	Fast Food	20 à 40 euros
7	21 à 40 ans	Indien	Moins de 20 euros
8	21 à 40 ans	Indien	Moins de 20 euros
9	21 à 40 ans	Indien	20 à 40 euros
10	21 à 40 ans	Etoilé	20 à 40 euros
11	21 à 40 ans	Etoilé	Plus de 40 euros
12	41 à 60 ans	Etoilé	Plus de 40 euros
13	41 à 60 ans	Indien	Plus de 40 euros
14	41 à 60 ans	Indien	20 à 40 euros
15	41 à 60 ans	Etoilé	20 à 40 euros
16	41 à 60 ans	Etoilé	Plus de 40 euros
17	Plus de 60 ans	Indien	20 à 40 euros
18	Plus de 60 ans	Etoilé	20 à 40 euros
19	Plus de 60 ans	Etoilé	Plus de 40 euros
20	Plus de 60 ans	Etoilé	Plus de 40 euros

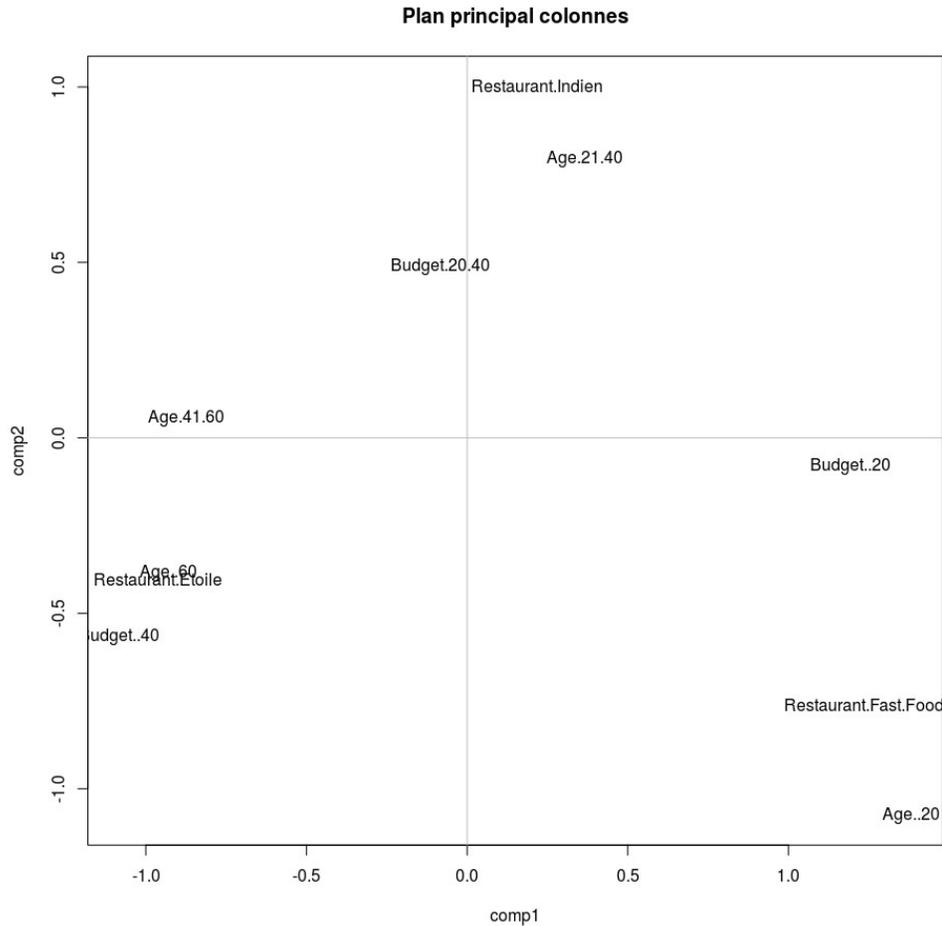
- Déterminer les tableaux de contingence des variables croisées deux à deux.
- Calculer le tableau de Burt B .

Ci-dessous et sur la page suivante on montre le résultat d'une ACM effectuée sur ce jeu de données.

- Discuter les inerties – combien d'axes factoriels devrait-on retenir ?
- Quelles sont les modalités les moins bien représentées sur le plan des deux premiers axes factoriels ? Que peut-on faire afin de mieux tenir compte de ces modalités ?
- Quelle est la variable qui explique le mieux le premier axe ? Et le deuxième axe ?
- Qu'est-ce qui caractérise le troisième axe ?

```
> acmin$TOT
      inertia      cum      cum(%)
Ax1 0.78077444 0.7807744 33.46176
Ax2 0.41325060 1.1940250 51.17250
Ax3 0.36283901 1.5568640 66.72274
Ax4 0.30641878 1.8632828 79.85498
Ax5 0.28385299 2.1471358 92.02011
Ax6 0.10405859 2.2511944 96.47976
Ax7 0.08213893 2.3333333 100.00000
```





Contributions absolues et relatives (multipliées par 10000):

```
> acmin$col.abs
```

	Axis1	Axis2	Axis3
Age.<20	1629	1872	40
Age.21-40	200	1784	104
Age.41-60	814	6	2947
Age.>60	740	238	2870
Restaurant.Etoile	1583	524	96
Restaurant.Fast.Food	1628	1166	184
Restaurant.Indien	71	2847	484
Budget.<20	1821	15	473
Budget.20-40	11	763	1945
Budget.>40	1498	780	851

```
> acmin$col.rel
```

	Axis1	Axis2	Axis3
Age.<20	4771	-2902	-55
Age.21-40	724	3402	174
Age.41-60	-2544	10	-4278
Age.>60	-2167	-369	3905
Restaurant.Etoile	-6180	-1084	175
Restaurant.Fast.Food	5084	-1927	268
Restaurant.Indien	256	5430	-811
Budget.<20	6094	-27	-735
Budget.20-40	-45	1578	3530
Budget.>40	-5015	-1382	-1324